DOCUMENT RESUME

ED 094 319 CS 001 188

AUTHOR Johnson, Dale D.
TITLE Word Lists That Make Sense--And Those That Don't.
PUB DATE May 74
NOTE 11p.; Paper presented at the Annual Meeting of the International Reading Association (19th, New Orleans, Louisiana, May 1-4, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Automatic Indexing; Beginning Reading; Elementary Education; Information Processing; *Reading Instruction; *Reading Research; Vocabulary; *Word Lists

ABSTRACT

Vocabulary studies conducted in this century are reviewed in this paper, with an emphasis on several recent investigations utilizing computer technology. The use of computers has greatly facilitated the ease and accuracy of word tabulation, but the lists are only as language-reflective as the sources from which they are derived. The great majority of vocabulary tabulations are derived exclusively from schoolbooks, a narrow source that tends to be self-perpetuating. Those lists developed from children's language, from frequency in general printed English, or from occurrence in literary and supplemental materials are considerably more relevant for text authors and, subsequently, for teachers. Because language changes, continual updating of word lists is necessary. (TO)

ED 094319

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Professor Dale D. Johnson
The University of Wisconsin
123 Education Building
Madison, Wisconsin 53706

"WORD LISTS THAT MAKE SENSE--
AND THOSE THAT DON'T"

by

Dale D. Johnson

The University of Wisconsin

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Dale D. Johnson
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

International Reading Association
New Orleans Convention

Friday, May 3
2:00-4:45 p.m.

SYMPOSIUM XXVII

CS 001 188

"Word Lists That Make Sense--And Those That Don't"

I wish to begin by stating that I still hold the somewhat old-fashioned conviction that written words are important in reading. I know it is more fashionable to be concerned with syntactic structures, semantic nuances and phonological relationships as important planks in bridging the gap from printed surface structure to the writer's or reader's deep structures. And I agree that they are important. Yet without words they are meaningless.

Syntactic structures--patterned, diagrammed, formulized or described--are useless without words: The formula "Article + Subject + Auxiliary + Verb + Article + Direct Object" is of no use to a reader until the words, "The boy can drive the car." have been inserted.

Similarly letter-sound correspondences, arrangements and sequences, be they labeled "rules," spelling patterns, decoding patterns, graphemic bases, phonograms, graphonemes or whatever, have no utility except in the context of words. We may wish to call words morphemes or free morphemes or "word-length units of meaning" (as one test does) but, however labeled, they are inescapably important components of language, which, in their written forms, must be dealt with by readers.

With this brief statement of bias as an introduction, I wish to spend my remaining minutes discussing things that I think are important (or unimportant) about word lists. For many decades reading teachers and researchers have been compiling lists of words they feel are useful for one purpose or another. For example, Cameron (1), while at a small college in Wisconsin, tabulated the profanity of undergraduate students as overheard in dormitories, hallways and campus taverns. In analyzing his results, Cameron neatly categorized such words

according to their derivation; sacred, excretory or sexual. From a different direction, Hill (6) compiled words found in best-selling comic books during World War II. Davis (4) prepared a list of what he termed "indispensible words" (such as bus stop, exit, toilet) common to everyday environment. Such lists are no doubt interesting and may, in some cases, be useful to young readers.

Obviously there are many purposes for and potential uses of word lists. Teachers of English as a second language may desire lists of words considered important to the oral language development of non-English speaking children. Researchers may require lists of CVC trigrams or tallys of homographs or homophones. Struggling textbook and test authors may wish for lists of picturable words or words common to a particular discipline. Spelling reformers like lists of words which demonstrate the peculiarities of English spelling while phonics advocates search for clusters of words that end in tch or ght or contain the /ɔ/ sound of o in medial position.

Our concern in this symposium is with vocabularies for beginning reading. We are asking the question "what kinds of words--that is which words-- do we consider important for young children to learn to read?" Certainly one's philosophy of what beginning reading should be determines the criteria believed to be important. In fact it is possible to neatly categorize the major approaches to beginning reading and many commercially produced reading series according to their beliefs about "first words"--the initial reading vocabulary.

Those among us who advocate organic reading, the language experience approach, would believe that the first words to be read should be those chosen by children--the words they want to learn to read. Those who believe in the importance of decoding in initial reading want to see "first words" which are consistent and patterned with regard to letter-sound correspondences. Proponents

of what we might collectively call "a basal reader" approach see the need for teaching high-frequency words--those words children will likely run into over and over again.

Nonetheless, many teachers of reading have, through the years, felt a need for what may be called "a basic sight word list." That is, teachers have desired a list of words considered vitally important for their pupils to know, regardless of underlying reading philosophy. And reading researchers have strived to meet these needs. In this decade alone many more than one hundred word lists--some very short and some very comprehensive--have appeared in the literature. Over 3,000 references are cited in the revised Bibliography of Vocabulary Studies by Dale, Razik and Petty (3). Some lists seem to have been more soundly derived than others. However, if we look at the scores of lists based on some sort of frequency count and look at, perhaps the one hundred or more most frequent words on each list, most of the lists look a lot alike.

I wish to suggest to you four postulates or canons, if you will, which I believe should guide the construction of word lists and the subsequent teaching, writing and research related to them.

1. The first is that no word list should be considered sacred, universally useful or final. Language changes constantly and new words enter a language daily.

Some words mean one thing today and another thing tomorrow. We could play a game we might call "Generation Gap." Just for a moment may I ask you to write some associations. As I read a word to you, jot down a word or two--perhaps a synonym or a definition--which the word brings to mind. (pot...salt... heavy...Apollo...bag...plumbers...) We could play this game all day.

Just as some words change in meaning, others fade in usage. Words such as shall, shaw, and gully, may have once been highly useful to learn, but are much less so today. A word list compiled in the 1920's may contain a number of words which are relatively less important in the 1970's. Lists compiled today will certainly need regular re-evaluation in the future.

2. My second canon of word list compilation is that such lists must be based on the language of children. Unless initial reading vocabularies contain words which are in the speaking and listening vocabularies of young children, they really cannot be very meaningful. Pioneer work such as that done by Horn (7) in 1926, the International Kindergarten Union in 1928 (8), Rinsland (12) in 1945 and more recent works such as those by Murphy (11) in 1957, Wepman and Hass (14) and Sherk (13) are studies generated from the oral speech of young children. These studies are quite massive, and because they present the words young children from various milieu use, should certainly be valuable sources for compiling the usually shorter sight-word lists used in reading.

3. A third canon is that beginning word lists, in addition to containing words known by children, should reflect the present-day world of printed American English in all its genre. By this I mean that the vast array of general printed matter as well as children's literature should be a basic source of sight-word vocabulary. Please note that I have not included basal readers.

In my opinion the dozens and dozens of studies compiling words found in basal reading series have been the most uninteresting and the most unproductive form of vocabulary research. There are two main problems with such studies. The reasoning behind them has often been illogical and certainly circular. Of what use is it to compile lists of 100 words or 400 words that are common to eight out of ten basic reading series? Is any child taught to read with this

wide array of basal series? Only in the wealthiest of school districts are more than two or three series purchased, and then they are usually intended for rather discrete groups of children. I have nothing against teaching children the new vocabulary they will encounter in their reading book; I think it is imperative, and I am happy that most basal series are carefully constructed to insure the learning of their vocabularies--through a variety of techniques. But it seems senseless to teach a word simply because it is found in six or eight controlled-vocabulary basal series.

There is a further criticism of vocabularies derived from basal series and that is their inherent circularity and stagnation. Some popular word lists published 30 or 40 or more years ago were derived from basic reading series then in use. Because of their popularity, such lists became vocabulary sources for a new generation of text authors. Then new lists were pulled from the new basals, and so it goes. I would much rather see the dog wag the tail than the tail wag the dog. It seems to me reading instructional materials written for children should contain the words the child will run into time and again in children's books and magazines and the broader world of printed English newspapers, books, magazines and the like.

We know that many children read much more widely than their school reading books and it seems that many other children would read more if they knew the words used in non-school materials. I argue that in addition to being taught the words in the reading series in use in the classroom, children should be taught a vocabulary of words they will frequently encounter elsewhere. And, of course, there will be overlap between the two.

Two recent studies, facilitated by the use of computer technology, have provided massive lists of words derived solely from textbooks written for

children. Harris and Jacobson (5) (1973) examined six basal reading series from grades one to six, and (commendably) also included two series each in social studies, English, math and science. They present a "core" list of words found in at least three of the six basal reading series. The recent computer-aided compilation by Carroll, Davies and Richman (2) (1971) is more useful in that it sampled magazines, novels, poetry and general non-fiction in addition to basic textbooks. However, it covered only materials intended for children in grades 3 through 9.

Three recent compilations provide very useful vocabulary sources for beginning word lists, I believe. They are: (1) The Kucera-Francis (9) (1967) study of 50,406 distinct words from more than one million running words found in five hundred 2,000-word samples drawn from 15 different genres including fiction, the sports page, etc. These words, particularly the top 500 or so, are the words most often found in printed American English. As such these top 500 or more words, particularly those that are also within the speaking-listening vocabularies of young children, would seem imperative words for teachers and reading textbook authors to utilize.

Another potentially valuable source from which basic sight words could be drawn is the compilation presented by Moe (10) (1973) based on his computer analysis of one hundred ten children's books--without controlled vocabularies--which were award winners or runners up in such contests as the Caldecott and Book World Children's Spring Festival. His list of 200 high frequency words accounted for 61% of the more than 100,000 running words. As with the top Kucera-Francis 500, these words should be a valuable source to teachers and authors.

Eighty popular children's library books were computer-analyzed by Durr (15) (1973) and of a total of more than 105,000 running words he presents the 188 words of highest frequency.

It seems that high frequency words from such studies as those done by Kucera-Francis, Moe and Durr when they also occur with high-frequency in the oral language of children, as identified by such studies as those by Murphy, Wepman and Hass, and Sherk, ought to be viewed as the currently most useful works from which to prepare sight word lists for teachers and authors.

4. A final point--really a side issue that needs more attention than can be given in our remaining minutes--concerns the meanings of words found in word lists. Too many sight word lists contain only the printed word without description of its function or meaning. For example, on Moe's list we see such words as can, saw, and head while on Durr's we find like, right and run. Using run as an example, we do not know if it equates to fast jogging, a hole in a stocking, water pouring from a tap, an attempt to be elected, a baseball score, or operating a business. Should we advise teachers to check a dictionary and teach the meaning listed first? Should we urge that all meanings for a word be taught? Or should we indicate the usage which we are presumably saying is so highly frequent? "My back is in back of my chest and I rarely take back what I've said about how poorly he backs up his car." But to quickly back off this issue I simply suggest that as we feed print into computers we should additionally provide sufficient instructions to the computer so that the resulting compilations tell us which words we really are advocating the use of.

In summary, word lists have been around a long time, and will continue to be. The use of computers has greatly facilitated the ease and accuracy of word tabulation--but the lists will only be as language-reflective as the sources

from which they are derived. Any list of basic sight words not derived from the language of children and high frequency in general printed English or children's literature beyond basal readers (with their controlled vocabularies) should be viewed suspiciously. Let's not put the cart before the horse. Let's let the language of children and the world of printed English dictate the reading vocabularies to be learned by children and to be found in instructional reading materials--rather than the reverse.

REFERENCES

1. Cameron, P. "The Language of College Students or Damn All Over," Unpublished Monograph, Stout State College, Menominee, Wisconsin, 1967, 6pp.
2. Carroll, J. B., Davies, P., and Richman, B. American Heritage Word Frequency Book, Boston: Houghton Mifflin, 1971.
3. Dale, E., Razik, T., and Petty, D. Bibliography of Vocabulary Studies, (5th Edition) Columbus, Ohio: Ohio State University, 1973.
4. Davis, D. C. "An Indispensible Sight-Word Vocabulary," Unpublished Monograph, The University of Wisconsin, Madison, 1969.
5. Harris, A. J. and Jacobson, M. D. Basic Elementary Reading Vocabularies, New York: MacMillan, 1972.
6. Hill, G. E. "The Vocabulary of Comic Strips," Journal of Educational Psychology, 34 (February, 1943) 77-87.
7. Horn, E. A Basic Writing Vocabulary, University of Iowa Monographs in Education, Series 1, No. 4, 1926.
8. International Kindergarten Union, Child Study Committee. A Study of the Vocabulary of Children Before Entering the First Grade, Washington, D.C.: International Kindergarten Union, 1928.
9. Kucera, N., and Francis, W. N. Computational Analysis of Present-Day American English, Providence: Brown University Press, 1967.
10. Moe, A. J. "Word Lists for Beginning Readers," Reading Improvement, Vol. 10, No. 2 (1973), 11-15.
11. Murphy, H., and others. "The Spontaneous Speaking Vocabulary of Children in Primary Grades," Journal of Education, Boston University, 140 (December, 1957), 3-106.

12. Rinsland, H. A Basic Vocabulary of Elementary School Children, New York: Macmillan, 1945.

13. Sherk, J. K. A Word Count of Spoken English of Culturally Disadvantaged Preschool and Elementary Pupils, Kansas City: University of Missouri, 1973.

14. Wepman, J. M., and Hass, W. A Spoken Word Count (Children - ages 5, 6, and 7), Chicago: Language Research Associates, 1969.

15. Durr, W. K. "Computer Study of High Frequency Words in Popular Trade Juveniles," The Reading Teacher, Vol. 27, (October, 1973) 37-42.